

openIDL - Architecture - Working Group Backlog

- [POC Scope](#)
- [Readiness for Production](#)
- [High Level Requirements](#)
- [Backlog](#)
- [Discussion](#)
 - [Data Verification](#)
 - [Data Validation](#)

Here we track the items in the backlog for architecture:

POC Scope

- current stat plan
- accepted architecture (as described here)
- can be a single upload
- for 2020 data
- validate against individual carrier data, only those that participate in the poc
- ?? reconciliation - quarter / annual
- ?? edit package

Success Criteria

- personal auto reports (2 identified) (for the participating carriers' data) matches
- manual integration between hosted node and carrier adaptor / hds

question

- is there a staging area in the hds

Readiness for Production

When is the architecture ready to be used in production?

- When all backlog items are adopted and implemented
- move data to published when requested (this leads to massive duplication)

High Level Requirements

- Privacy for Data Owners
 - carrier raw data is private
- Data Quality for the Consumers
- able to participate in the process
- auditable participation in the process
- carrier raw data is in a common format
- data flow to final report must be secured

Backlog

	Description	Decisions	Adoption Status	Date of Status	Spike /POC
Data Verification	How do we verify the data is available for a data call / regulatory report? This is not a measure of its quality (see below)	<ul style="list-style-type: none">• The verification that the data is available is made at the time of the extraction by the consenter.• The required data format/level/version is specified on the extraction. This is effectively the requirement for the data.• By consenting to the extraction, the data owner asserts that the data meets the format/level/version requirements.• (Perhaps hold the result of the extraction for some period of time)			

Data Quality Validation	What validation of the data is provided? Is it all part of openIDL? Is it shared across the community? Is there some provided by the carrier/member?	<ul style="list-style-type: none"> For stat reporting, the stat-agent is required to validate the data for accuracy. The validation rules are common and can be shared across all data owners. 			
Edit Package	SDMA provides a way to identify and fix errors as they occur before submission of the data. Does this belong in the openIDL as a community component? If so, how do we provide that?	<ul style="list-style-type: none"> A reference implementation is made available for the etl that updates the HDS The data owner is responsible for the implementation of the ETL The data must meet certain levels of correctness (see SDMA for guidance) 			
Adapter Hosting Approach	The adapter runs multiple components required by participants in openIDL. How can we host these in member environments?	<ul style="list-style-type: none"> Software running in the Adapter is provided as Docker containers The images are maintained in open source. Members are free to vet and bring the images internally. A mechanism for updates must be agreed upon. This includes an SLA that describes the timeframe for absorbing updates. 			
Adapter Components	What are the components in the Adapter?				
Extraction Technology	What is the technology used to execute the extractions in the member environment?				
Data Standard Format	What is the format for the data? The data at rest in the HDS is the same or different from the data as it is being validated?	<ul style="list-style-type: none"> Every format has a name Every format optionally has a level Every format / level has a version Every format level version has a document or section of a document that describes it <ul style="list-style-type: none"> The fields are enumerated Each field includes enough information on how to populate it The standard for verification <ul style="list-style-type: none"> Timeliness Completeness The standard for quality <ul style="list-style-type: none"> Valid values for all fields Percent of acceptable errors Issue levels for each possible validation The rules needed (in prose and/or in technology) to validate the data. 			
Harmonized Data Store and History	What history is maintained in the harmonized data store?	<ul style="list-style-type: none"> Full history of transactional change Snapshots that roll up the transaction Do we prescribe the transaction model? Need a staging database? Is this part of the openidl spec or only responsibility of the carrier. Do we agree that the HDS follows a "Data Vault" model? https://en.wikipedia.org/wiki/Data_vault_modeling 			

Discussion

Data Verification

When data is inserted into the harmonized data store, the insertion is registered on ledger as passing all validation and supporting some level of extraction.

Is there a need to capture a hash of the data?

The data inserted into the HDS is captured per policy. The history of the transactions is captured, so the "state" of the policy at a given date may change between the time data is initially inserted and when it is extracted.

The data extraction will state what level of data is required. How does the member attest that the data meets that expectation? Does the consent constitute that attestation?

Data Validation